

Estimating gray whale abundance from shore-based counts using a multilevel Bayesian model

DURBAN, J.W., WELLER, D.W., LANG, A.R., PERRYMAN, W.L.

Marine Mammal and Turtle Division, Southwest Fisheries Science Center, National Marine Fisheries Service,
National Oceanic and Atmospheric Administration, 8901 La Jolla Shores Dr., La Jolla, CA 92037, USA
Contact email: john.durban@noaa.gov

ABSTRACT

Counts of southbound migrating whales off California form the basis of abundance estimation for the eastern North Pacific stock of gray whales (*Eschrichtius robustus*). Previous assessments (1967-2007) have estimated detection probability (p) from the detection-non detection of pods by two independent observers. However, tracking distinct pods in the field can be difficult for single observers; resulting in biased estimates of pod sizes that needed correcting, and matching observations of the same pod by both observers involved key assumptions. Due to these limitations, a new observation approach has been adopted wherein a paired team of observers work together and use a computerized mapping application to better track and enumerate distinct pods and tally the number of whales passing during watch periods. This approach has produced consistent counts over four recently monitored migrations (2006/7, 2007/8, 2009/10 and 2010/11), with an apparent increase in p compared to the previous method. To evaluate p and estimate abundance in these four years, we compared counts from two independent stations of paired observers operating simultaneously using a hierarchical Bayesian “ N -mixture” model to jointly estimate p and abundance without the challenge of matching pods between stations. The baseline detectability p_o was estimated as 0.80 (95% Highest Posterior Density Interval [HPDI] = 0.75-0.85), which varied with observation conditions, observer effects and changes in whale abundance during the migration. Abundance changes were described using Bayesian model selection between a parametric model for a Normally distributed common migration trend and a semi-parametric model that estimated the time trends independently for each year; the resultant migration curve was a weighted compromise between models, allowing for key departures from the common trend. The summed estimates of migration abundance ranged from 17,820 (95% HPDI = 16,150-19,920) in 2007/8 to 21,210 (95% HPDI = 19,420-23,230) in 2009/10, consistent with previous estimates and indicative of a stable population.

KEYWORDS Abundance estimate, Migration, Modeling.

INTRODUCTION

The eastern North Pacific stock of gray whales migrates annually along the west coast of North America from high latitude feeding grounds to winter breeding grounds in the lagoons and adjacent ocean areas off Baja California, Mexico (Rugh *et al.* 2001). This nearshore migration pattern has enabled repeated abundance estimates from shore-based counts off Granite Canyon, central California. In 23 years, between 1967 and 2007, counts of the number of observed pods travelling southbound have been rescaled using estimates of pods undetected during watch periods, pods passing outside watch periods, and night travel rate (Buckland *et al.*, 1993; Laake *et al.*, 1994; Buckland and Breiwick, 2002; Hobbs *et al.*, 2004; Rugh *et al.*, 2005; Laake *et al.*, 2012). Population models based on these estimates indicate that gray whales have increased substantially in population size, recovering from whaling operations in the 19th and 20th centuries, and are now close to carrying capacity and likely pre-exploitation levels (Punt and Wade, 2010). The most recent population estimate from abundance counts in 2006/2007 was approximately 19,000 whales (Laake *et al.*, 2012).

To facilitate continued population monitoring, the abundance estimation approach has seen continual evolution throughout this time series to more realistically estimate detection probability (p) to link observed counts to true abundance; this paper describes the latest modification. Notably, previous assessments have estimated p from the detection-non detection of pods by independent observers using an analytical mark-recapture approach. However, tracking distinct pods in the field is difficult (Rugh *et al.* 2008), particularly for a single observer using just hand-recorded entries onto a paper data form. As a result, matching observations of the same pod by both observers

involved key (and untestable) assumptions, and limited observations of a given pod required corrections for bias in pod size estimation (Rugh et al. 2008; Laake *et al.*, 2012). Due to these limitations, a new observation approach has been developed wherein a paired team of observers work together and use a computerized mapping application to help better track distinct pods and tally the number of whales passing during watch periods (Durban *et al.*, 2010). This approach has a number of advantages, including open communication between observers, enabling one observer to search for whales continually without the distraction of looking down to record data, and easier separation and tracking of distinct pods due to the precise computerized data recording and visualization. As a result, this approach enables more repeated observations of each pod, leading to larger (and presumably less biased) estimates of pod size (Durban *et al.*, 2010), and has produced consistent counts over four recently monitored migrations (2006/7, 2007/8, 2009/10 and 2010/11), with an apparent increase in p compared to the previous method (Durban *et al.*, 2011).

To evaluate p for this new approach, here we compared watch period counts from two independent stations of paired observers operating simultaneously during two of the four years (2009/10 and 2010/11), using a hierarchical Bayesian “ N -mixture” model (Royle, 2004) to jointly estimate the probability of detection and abundance in all four years, without the challenge of matching pods between stations. This “ N -mixture” approach has been successfully used to estimate abundance and detectability from replicate count data for a range of wildlife species where it has not been possible to match repeat sightings of individuals (e.g. Kery *et al.*, 2005; Joseph *et al.*, 2009; Chelgren *et al.*, 2011). Here we demonstrate the utility of this approach to extend the time series of abundance estimates for eastern North Pacific gray whales.

METHODS

Data Samples

Counts of gray whales were conducted from shore-based watch stations at Granite Canyon, California, during the 2006/7, 2007/8, 2009/10 and 2010/11 southbound migrations (Table 1). Counts were made by rotating teams of observer pairs using naked eye aided by 7x50 binoculars; the primary observer in the pair kept continual visual watch while the secondary observer served as a data recorder but also kept watch and assisted with tracking already identified pods whenever possible. Each observer had one 90 min shift as primary observer, followed by a second 90 min shift as secondary observer, and then a 90 min break. Sightings were entered into a real-time data logging PC program, which had a mapping screen to help track repeated sightings of the same pod. The map projected the likely movement tracks (and error ellipses) of the pods using predicted swimming speeds ($1.44 - 1.95 \text{ ms}^{-1}$), allowing re-sightings and new sightings to be queried. Up to six 1.5-hour watch periods were used to cover daylight hours from 07:30 to 16:30 local time, during which the observers recorded passing whales and environmental conditions, specifically visibility (subjectively categorized from 1 to 6 for excellent to useless) and sea state (Beaufort scale). To control for weather conditions and for consistency with previous abundance estimations, we only used counts during watch periods with acceptable weather conditions throughout their entire duration, specifically visibility code <5 (excellent to fair) and Beaufort Scale <5.

Table 1: The number of whales recorded during the southbound gray whale surveys from 2006/7 to 2010/2011. Data are the total counts, hours and distinct days for watches during acceptable observation conditions.

Migration	North Station			South Station		
	Dates	Whales	Hours (days)	Dates	Whales	Hours (days)
2006/7	02 Jan–03 Feb	2691	204 (34)	-	-	-
2007/8	02 Jan–09 Feb	2079	202.5 (34)	-	-	-
2009/10	30 Dec–19 Feb	2034	246 (43)	11 Jan–06 Feb	1551	105 (20)
2010/11	03 Jan–18 Feb	2885	265 (45)	10 Jan–04 Feb	1754	141 (24)

Estimating Detection Probability

We used the “ N -mixture” approach (Royle, 2004) to simultaneously estimate detection probability p_{ijt} and abundance N_{jt} for each watch period j in each year t , based on the total aggregated counts n_{ijt} of passing whales recorded by each of $i = 1:2$ watch stations in each period. The observed counts n_{ijt} were modeled as a binomial outcome conditional on the unknown true number of whales passing N_{jt} and the detection probability p_{ijt} with hierarchical models assumed to describe variability in both N and p (e.g. Chelgren et al. 2011). The power to estimate detectability was achieved by comparing gray whale counts from two independent stations of paired observers operating simultaneously during two years (2009/10 and 2010/11) from watch stations that were positioned 35m apart at the same elevation (22.5m) above sea level. In 2009/10 counts were compared from the two watch stations operating simultaneously during 70 1.5-hour watch periods with acceptable weather conditions, covering 20 different days of the migration; in 2010/11 simultaneous counts were available from 94 watch periods over 24 different days (Table 1). However, we could also extrapolate detectability for all monitored watch periods in each of the four years based on the fitted model for detectability. In order to accomplish this, the counts for the south watch station were treated as zero-inflated binomial outcomes, with the binomial probability specified as a function $u_{ijt} p_{ijt}$ where $u = 1$ or 0 to indicate whether or not count data were actually collected from that station, thus ensuring that structural zero counts from periods without a second watch did not contribute to the likelihood for estimation of p or N .

Consistent with Laake *et al.* (2012), the model for detectability incorporated fixed effects β for visibility (VS) and Beaufort Scale (BF), as well as random effects associated with each observer o in 1:OB observers. These were modeled as additive effects on a general intercept so that the direction and magnitude of the estimated effects away from zero (no effect) could be assessed. We selected for the inclusion of these effects using Bayesian model selection with stochastic binary indicator variables g to switch each of the three possible effects either in or out of the model depending on their relevance to the observed data (Kuo and Mallick, 1998):

$$\text{logit}(p_{ijt}) = \text{logit}(p_o) + g^{bf} \beta^{bf} \text{BF}_{jt} + g^{vs} \beta^{vs} \text{VS}_{jt} + g^{ob} \beta_{ijt=o}^{ob}$$

where the intercept p_o was the base detection probability in the absence of covariate effects, assigned a Uniform(0,1) prior distribution, and $\text{logit}(p_o) = \ln(p_o/1-p_o)$. Centered around this base detectability, each of the fixed effects β^{bf} and β^{vs} was assigned a Normal prior distribution with mean zero and large standard deviation of 10; this prior value was smaller than the corresponding posterior estimates of standard deviation, and as such this was vague prior distribution that allowed any non-zero effects to emerge if supported. The random effect for each observer $\beta_{o=1:n.obs}^{ob}$ was drawn from a Normal distribution with mean zero and standard deviation $\sigma^{ob} \sim \text{Uniform}(0,10)$. Each of the binary indicator variables, g , was assigned a Bernoulli(0.5) distribution to specify equal probability of inclusion or not of the effect in the model.

Fitting Migration Curves

The N -mixture approach also accounted for variation in p relative to changes in N (latent watch period abundances) during the migration. So, some power to estimate detectability was achieved by assuming a model for changes in watch period abundance over the course of the migration. We adopted a Poisson distribution $N_{jt} \sim \text{Poisson}(\lambda_{jt})$ as a hierarchical prior for the distribution of abundances, and specified a model for the Poisson mean λ in terms of the number of whales passing each day (d), with an offset for the effort duration of each watch period, E_{jt} in decimal days (e.g. Laake *et al.*, 2012),:

$$\log(\lambda_{jt}) = \log(E_{jt}) + \text{model}_{d(j)t}$$

$$\text{model}_{dt} = z_{dt} \text{Common}_{dt} + (1-z_{dt}) \text{Specific}_{dt}$$

Days were specified as $d = 0$ to D_t . In all four years t we used $D_t = 90$, where days were counted from 12:00am on 1 December, and we added an abundance of 0 whales passing for day 0 and D_t to anchor the fitted model when we assumed whales did not pass (following Buckland et al. 1993). Estimates from the remaining days were derived from a mixture (or compromise) of two competing models (“Common” and “Specific”; e.g. Li *et al.*, 2012) describing changes in abundance across each annual migration. The model contributing each daily estimate was indicated using stochastic binary indicator variables z_{dt} , each assigned a non-informative Bernoulli(0.5) prior distribution. As such, each z_{dt} indicated the probability of a daily estimate conforming to the common trend,

allowing flexibility for departures from this trend that may only exist on certain days in certain years to be identified and modeled (rather than assuming all counts from an entire year conform to or depart from a common trend, which would be represented by z_t). The total number of whales passing during each migration was then estimated by summing the expected value from the model-averaged number of whales passing each day from time 0 to D_t (e.g. Laake *et al.*, 2012). These estimates were then rescaled to account for the differential passage rate at night (Perryman *et al.*, 1999), based on the 9-hour day multiplicative correction factor of Rugh *et al.* (2005). Specifically, we applied a constant nighttime correction factor that was assumed to be a Normally distributed fixed effect with mean of 1.0875 and standard deviation of 0.037.

For the “Common model”, we assumed a typical trend in abundance throughout each annual migration (e.g. Buckland *et al.*, 1993), with abundance changes assumed Normally distributed around a migration mid-point, with a Normal distribution specified as a quadratic function of days, on the log scale:

$$\text{Common}_{dt} = a_t + b_t * d_t + c_t * d_t^2$$

where the mid-point of the migration curve for each year t was derived by $-b_t/2a_t$. This assumed common migration curve allowed information to be “borrowed” across years when needed, specifying association across years to strengthen inference about migration curves in years with relatively sparse counts. However, we specified each of the curve parameters a_t , b_t and c_t to be drawn from hierarchical Normal distributions with means $\mu^a, \mu^b, \mu^c \sim N(0, 10)$ and standard deviations $\sigma^a, \sigma^b, \sigma^c \sim \text{Uniform}(0,10)$; hyper-parameters that were common across years, rather than assuming that the parameters themselves were constant. This random effects formulation allowed the timing, level and extent of the Normal migration curve to vary annually around the general pattern, if supported by the data.

Although it is likely that there is a typical pattern to the migration, we acknowledged that abrupt departures may occur at any time in any particular year. To incorporate unusual patterns, we allowed for the selection of an alternative “Specific” migration model: a semi-parametric model that estimated the time trends independently for each year (e.g. Laake *et al.*, 2012). We adopted a method in which the shape of the relationship of abundance across days was determined by the data without making any prior assumptions about its form, by using penalized splines (Ruppert, 2002). Following Crainiceanu *et al.* (2005) we used a linear (on the log scale) penalized spline to describe this relationship:

$$\text{Specific}_{dt} = S_{0t} + S_{1t} * d_t + \sum_{k=1}^m \lambda_{kt} (d_t - \kappa_{kt})$$

Where $S_{0t}, S_{1t}, \lambda_{1t}, \dots, \lambda_{kt}$ were regression coefficients to be estimated separately for each year and $\kappa_{1t} < \kappa_{2t} < \dots < \kappa_{kt}$ were fixed knots. We used $m = 15$ knots; a relatively large number to ensure the desired flexibility, and let κ_{kt} be the sample quantile of d_t 's corresponding to probability $k/(m+1)$. To avoid overfitting, we penalized the λ 's by assuming that these coefficients of $(d_t - \kappa_k)$ were Normally distributed random variables with mean 0 and standard deviation $\sigma_t^y \sim \text{Uniform}(0,10)$. The parameters S_{0t}, S_{1t} were modeled as fixed effects with Normal(0, 10) prior distributions.

Bayesian Inference using MCMC

The multi-level model was fit using Markov Chain Monte Carlo (MCMC) sampling using the WinBUGS software (Lunn *et al.*, 2000). Inference was based on 15,000 repeated draws from the posterior distribution of each model parameter conditional on the observed data, following 5000 iterations that were discarded as burn-in. Convergence of parameters within these initial 5000 iterations was determined based on Gelman-Rubin statistics below 1.05 (Brooks and Gelman, 1998) calculated from three independent chains begun from over-dispersed starting values. To gauge the adequacy of the model for each annual set of count data, we computed Bayesian P-values (Gelman *et al.* 1996) by using the same MCMC sampler to predict a distribution for each watch-period count from the posterior estimates of model parameters and comparing the total predicted and observed counts. For each year, there was good agreement between the model predictions and observed counts, with Bayesian P-values ranging from 0.45 to 0.53; values close to 0.5 would indicate that the data was consistent with replications under the model, with the distribution of the predicted count symmetrically overlapping the observed count (Gelman *et al.* 1996).

The MCMC sampling approach allowed uncertainty to be propagated across levels of the model. Notably, estimates of parameter values across MCMC iterations were used to estimate the probability of inclusion of covariate effects

in the model for detectability, given by the posterior probability $p(g=1)$ of each indicator variable g . Fitting and selection of the two competing migration models was achieved within the same MCMC run using the “cut” function in WinBUGS to ensure that estimation of the two models was not affected by the selection of the model indicator (e.g. Li et al. 2012). The posterior probability of conforming to the common trend model was then calculated by the relative frequency that each model was selected by the indicator z_{dt} in the overarching mixture model, and inference about abundance on each day was based on a weighted compromise between the competing models by sampling across the posterior distribution of z_{dt} .

RESULTS

The base detectability was estimated as $p_o=0.80$ (95% Highest Posterior Density Interval [HPDI] =0.75-0.85), which was modified by observation conditions and observer effects (Table 2). The posterior distribution for the effect of sea state β^{bf} , measured using the Beaufort scale, largely overlapped with zero and there was therefore low support for including this effect in the model with $p(g^{bf}=1)=0.004$. In contrast, there was a relatively strong negative effect of visibility on detectability (higher visibility code=lower visibility=lower detectability), with the entire distribution for β^{vs} falling below zero [$p(g^{vs}=1)=1$]. There was also support for inclusion of observer effects [$p(g^{bs}=1)=1$], with both positive and negative effects reflecting relatively high and low counts by different observers. A total of 35 different observers were used over 4 years between North and South stations; 15/35 counted in multiple years (2 years = 7, 3 years =4, 4 years=4). The Posterior medians for observers’ effects ranged from -0.59 to 0.80, but only five observer effects (all positive) had posterior distributions that did not include zero.

Table 2: Parameters of models for detectability, p . All estimates are presented as the 2.5%, 50%, 97.5% highest density posterior probability intervals, plus the probability of inclusion in a model (if tested), given by the posterior probability $p(g=1)$ of each indicator variable g . Observers are arbitrarily numbered, differently for each year.

	2006/07	2007/08	2009/10	2010/11
<i>Detection model</i>				
p_o	0.75, 0.80 , 0.85	0.75, 0.80 , 0.85	0.75, 0.80 , 0.85	0.75, 0.80 , 0.85
β^{bf} [$p(g^{bf}=1)$]	-19.34, -0.003 , 19.98 [0.004]	-19.34, -0.003 , 19.98 [0.004]	-19.34, -0.003 , 19.98 [0.004]	-19.34, -0.003 , 19.98 [0.004]
β^{vs} [$p(g^{vs}=1)$]	-0.38, -0.30 , -0.20 [1]	-0.38, -0.30 , -0.20 [1]	-0.38, -0.30 , -0.20 [1]	-0.38, -0.30 , -0.20 [1]
σ^{ob} [$p(g^{bs}=1)$]	0.26, 0.37 , 0.54 [1]	0.26, 0.37 , 0.54 [1]	0.26, 0.37 , 0.54 [1]	0.26, 0.37 , 0.54 [1]
Observer 1	-0.36, 0.02 , 0.49	0.03, 0.37 , 0.81	-0.42, -0.24 , 0.06	-0.13, 0.08 , 0.30
Observer 2	0.03, 0.37 , 0.81	-0.78, -0.03 , 0.70	-0.09, 0.30 , 0.81	-0.36, 0.02 , 0.46
Observer 3	-0.24, -0.07 , 0.11	-0.24, -0.07 , 0.11	0.03, 0.37 , 0.81	-0.42, -0.24 , 0.06
Observer 4	-0.42, -0.01 , 0.49	-0.42, -0.24 , 0.06	-0.13, 0.08 , 0.30	-0.25, 0.01 , 0.29
Observer 5	-0.04, 0.14 , 0.35	-0.13, 0.08 , 0.30	-0.24, -0.07 , 0.11	0.16, 0.43 , 0.73
Observer 6	0.06, 0.42 , 0.83	-0.04, 0.14 , 0.35	-0.27, -0.06 , 0.18	-0.04, 0.14 , 0.35
Observer 7	-0.17, 0.11 , 0.46	-0.18, 0.19 , 0.61	-0.04, 0.14 , 0.35	-0.50, -0.13 , 0.26
Observer 8	-0.39, -0.16 , 0.07	-0.17, 0.11 , 0.46	0.12, 0.33 , 0.59	-0.39, -0.16 , 0.07
Observer 9	0.12, 0.33 , 0.59	0.12, 0.33 , 0.59	-0.25, 0.01 , 0.29	-0.09, 0.23 , 0.60
Observer 10	-	-0.39, -0.16 , 0.07	-0.08, 0.26, 0.64	-0.27, -0.06 , 0.18
Observer 11	-	-	-0.71, -0.43 , 0.13	0.31, 0.80 , 1.46
Observer 12	-	-	-0.66, -0.37 , 0.07	-0.54, -0.29 , 0.04
Observer 13	-	-	-0.42, 0.00 , 0.49	-0.75, -0.22 , 0.33
Observer 14	-	-	-0.63, -0.13 , 0.40	0.12, 0.33 , 0.59
Observer 15	-	-	0.31, 0.80 , 1.46	-0.73, -0.29 , 0.14
Observer 16	-	-	-0.18, 0.19, 0.61	-0.18, 0.19 , 0.61
Observer 17	-	-	0.16, 0.43 , 0.72	-0.70, 0.02 , 0.76
Observer 18	-	-	-0.39, -0.16 , 0.07	-0.63, -0.13 , 0.40
Observer 19	-	-	-0.22, 0.22 , 0.72	-0.83, -0.59 , 0.36
Observer 20	-	-	-0.28, 0.14 , 0.59	-0.24, -0.07 , 0.11
Observer 21	-	-	-0.18, 0.28 , 0.83	-0.21, 0.11 , 0.47
Observer 22	-	-	-	-1.05, -0.49 , 0.06

Detectability also varied with changes in whale abundance during the migration, as shown by the extent of extrapolation from the daily summed counts (effort adjusted) to the estimated daily abundances (Figure 1). Detectability declined with increasing abundance, with a greater proportion of whales estimated to be missed as more whales passed during busy watch periods. In general, changes in abundance during the migrations were adequately described by a Normal curve over time, but there was greater uncertainty in the tails of the distribution resulting from generally sparse coverage. The Normal trend was useful for comparing migration timing: the median of the curve midpoints was 53.5 days since December 01 (23-24th of January), ranging between 49 and 57 days. However, there were some notable deviations from the Normal trend, with estimates from the year-specific non-parametric trend model being favored for some days in each of the four years. In particular, there was a high probability in favor of the Specific model [$p(z=0)>0.75$] on 9 days in 2006/7, 9 days in 2007/8, 16 days in 2009/10 and 11 days in 2010/11, representing key departures from the Normal migration trend. The summed (model-averaged) estimates of migration abundance ranged from a posterior median of 17,820 (95% HPDI = 16,150-19,920) in 2007/8 to 21,210 (95% HPDI = 19,420-23,230) in 2009/10, consistent with previous estimates (Figure 2). These new estimates were also relatively precise with coefficients of variation (CV = Posterior Standard Deviation / Posterior Median) ranging from 0.04 to 0.06 (median= 0.05), but nonetheless the 95% HDPI's of all four estimates overlapped.

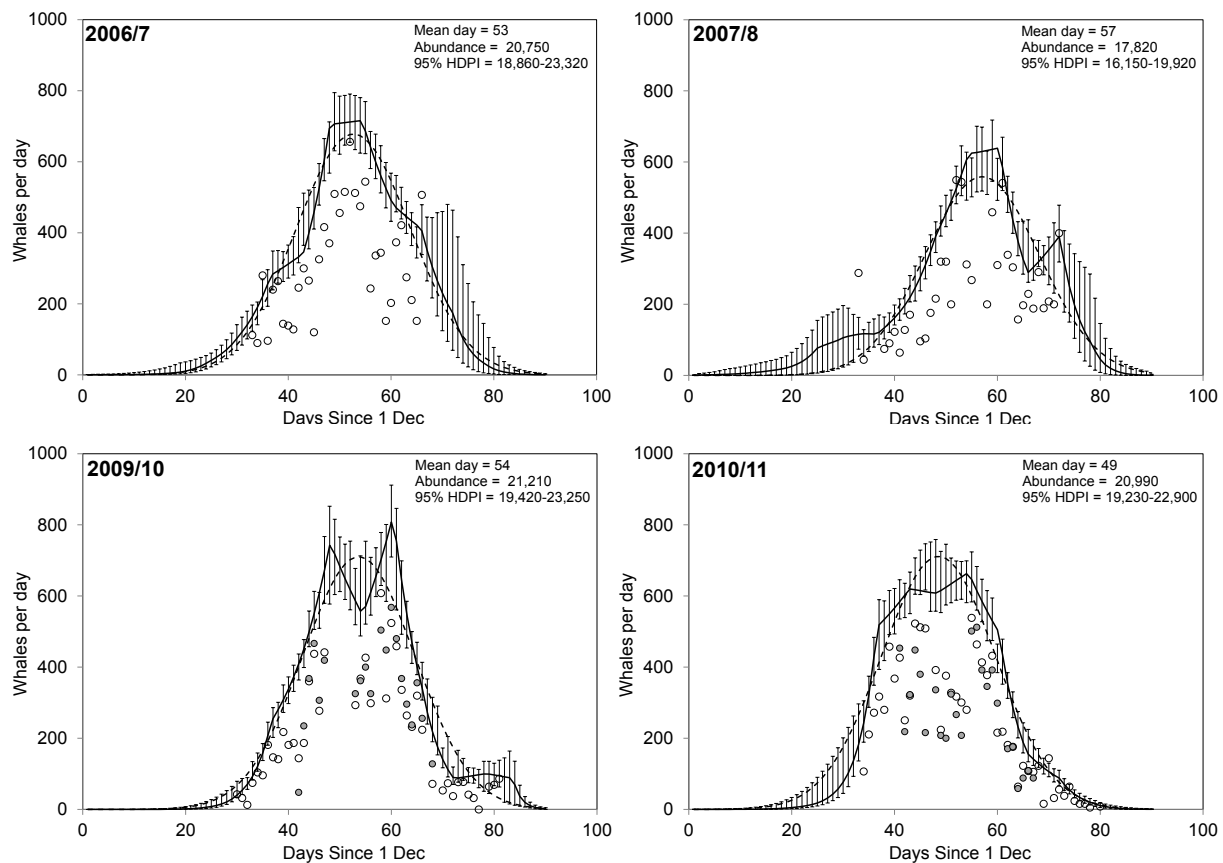


Figure 1. Observed whale passage rates expressed as total counts per day/ proportion of day observed (circles) and fitted migrations models (lines) for the four southbound gray whale migration counts from 2006/7 to 2010/11. Solid circles represent counts from a second watch station, when operating. The broken line represents the median estimates from a hierarchical Normal model for migration and the solid line represents a semi-parametric model of penalized splines; the abundance estimate for each day (95% highest posterior density interval shown by vertical lines) is a model averaged compromise between the migration models, and these were summed to estimate the overall abundance for the migrations.

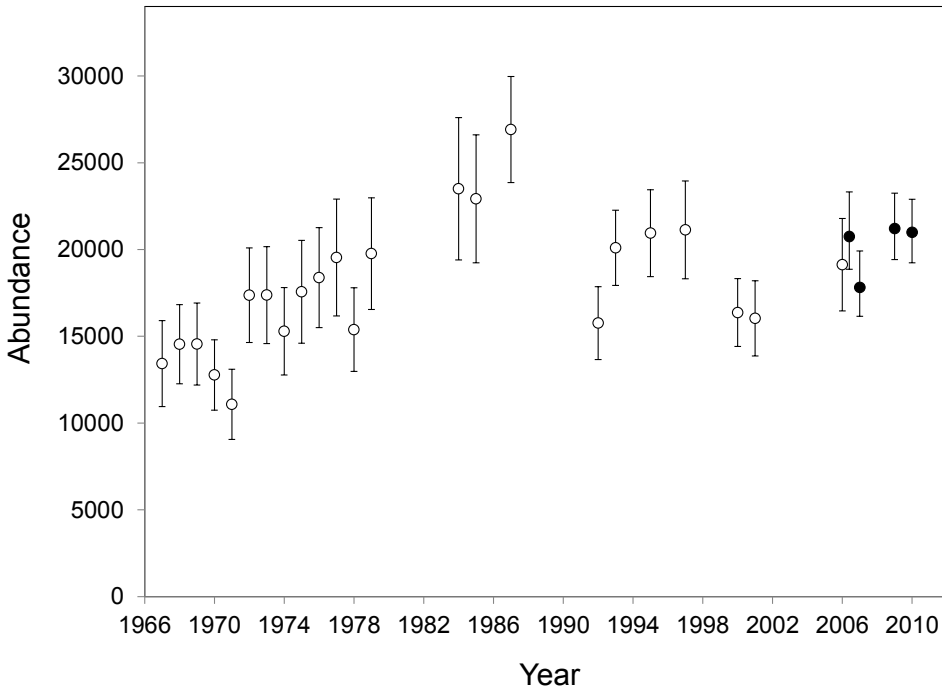


Figure 2: Gray whale abundance estimates for each of 23 southbound migrations with an end year between 1967 and 2007 (open circles, with 95% confidence intervals; from Laake *et al.*, 2012) together with the four recent migrations reported here (closed circles show posterior medians, lines are 95% highest posterior density intervals).

DISCUSSION

The new counting method adopted here was intended to reduce reliance on the ability of single observers acting independently to record and track distinct whale groups. By adopting teams of paired observers working together, with the benefit of a real-time computerized tracking and visualization tool, this approach has proved successful in increasing detection probability (Durban *et al.*, 2011) and also reducing variability in detections due to observer effects. Although still present, the magnitude of observer effects estimated from the new counts (Table 2) was generally not as great as those apparent with the traditional counting approach (see Laake *et al.*, 2012, Table 7). Furthermore, our method for estimating detectability departed from the mark-recapture approach of matching detections and non-detections of specific pods by independent observers. Instead we based inference on total watch period counts that were not sensitive to differential lumping and splitting of pods by observers, and avoided the assumptions required to match observed pods between pairs of observers. As an alternative to the mark-recapture analytic approach, we have shown how tallied watch period counts from two observer pairs counting simultaneously can lead to similar inference when analyzed using the N -mixture approach (Royle, 2004).

The N -mixture approach is conceptually simple: multiple observations of watch period counts, n , from the different observer teams represented different samples from an unknown binomial distribution with total population size N and detection probability p . A binomial likelihood function could then be easily used to estimate N and p from the sample of n 's. Although there were only a maximum of two samples of n during any specific watch period, a large sample of n 's was built up across many watch periods, allowing the estimation of the parameters. Layered on top of this core estimation process were both a trend model for true daily abundance through time based on the migration pattern and a model for how detection varied according to environmental conditions and different observers. Specifically, a hierarchical model fit to the replicate count samples allowed us to link detectability to key covariates, as in previous gray whale assessments (e.g. Laake *et al.*, 2012), and also extrapolate detectability based on these covariate relationships for watch periods without replicate counts. Similarly, by assuming a common underlying model for the migration pattern, this approach notably accounted for variation in p relative to changes in abundance N during the migration. Furthermore, this joint modeling of data from multiple years allowed the borrowing of

strength across years to better parameterize the migration during years with sparse data. Previously, two contrasting approaches have been used to model changes in abundance over the course of the annual gray whale migration: either by assuming a parametric model to determine the shape of the migration curve (Buckland *et al.*, 1993) or by fitting a non-parametric smoother to allow the data to determine the trend in abundance over time (Laake *et al.*, 2012). Here we drew on elements of both these approaches in a flexible framework using Bayesian model selection between a parametric model for a common migration trend and a semi-parametric model that estimated the time trends independently for each year; the resultant migration curve was a weighted compromise between models, allowing for key departures from the common trend.

The abundance estimates produced for 2006/7, 2007/8, 2009/10 and 2010/11 were internally consistent, consistent with previous estimates and indicative of a stable population (Figure 2). The 95% HDPI's of all four estimates overlapped, and there was substantial overlap between the 95% HDPI from the 2006/7 estimate with the 95% confidence intervals of the estimate for the same migration produced using the previous counting and estimation approach (Laake *et al.* 2012). Further, our estimates are very similar to the predictions of Punt and Wade (2012) based on assessment models for the full time series; their baseline model prediction for 2009/10 had 90% posterior density intervals ranging from 17,726 to 23,247; the posterior distribution for our 2009/10 estimate was centered within these intervals at 21,210 (95% HPDI = 19,420 – 23,250). It is noteworthy that the estimates produced using our approach were relatively precise with CVs ranging from 0.04 to 0.06 (median = 0.05) in contrast to CVs ranging from 0.06 to 0.09 (median = 0.08) for the 23 previous estimates.

This consistency provides a level of confidence in our approach and resultant estimates, but nonetheless there are limitations to address. Our approach makes a number of important modeling assumptions, both in terms of distributional forms and model structure. Notably we assumed that the detectability relationships described by modeling repeated counts during two years were also applicable in the remaining two years with no replicate counts. We also assumed observer effects remained constant, although in reality this may change with experience. Additionally, the definition of what constituted the common migration trend was dependent on the joint modeling of just four years of data, and precise inference about the shape of the migration curve relies on count data being collected from throughout the migration time span. During at least 3 / 4 years reported here, count data were sparse (or non-existent) during the tails of the migration, resulting in uncertainty over the shape of the abundance curve. While this uncertainty was propagated into inference about overall abundance in our Bayesian inference using MCMC sampling, the resulting imprecision will ultimately constrain power to detect between-year changes in migration patterns and abundance. Data collected during further migrations will be incorporated into this hierarchical model and therefore used to refine parameter estimates; this will benefit from replicate counting experiments, repeated when possible. As the time series grows, specific goodness-of-fit tests should be adopted to investigate aspects of model structure and suggested changes as necessary.

There are also practical considerations as well as modeling assumptions. Previous work has shown that the new counting approach produces estimates of pod size that are typically larger (and presumably less biased) than the traditional counting approach (Durban *et al.*, 2010), likely because the computerized tracking software facilitates more repeated observations of the same groups. In fact, we have assumed here that estimates of pod size using this observation approach are effectively unbiased and have not been rescaled to tally watch period counts. This is an assumption that remains to be tested, but suitable calibration experiments are difficult to design and implement, particularly due to the inherently subjective differences between observers in lumping and splitting whales to define groups. Similarly, although observer effects have been accommodated in the model for detectability, it is clear that too many observers (35 in total) counted too infrequently to allow precise parameterization of their relative effects on detectability in many cases. This will have resulted in further imprecision.

Although there may be field protocols that could be adapted to address these limitations within the current approach, we recommend further modernization of the observation process. Specifically, more accurate information could be gleaned from observations recorded with high-definition video files to allow subsequent review and re-review, rather than relying on instantaneous assessment by visual observers. The use of infra-red sensors would further allow for 24-hour monitoring (e.g. Perryman *et al.*, 1999) and provide greater coverage of the entire migration during acceptable weather conditions; automated blow detectors (e.g. Santhaseelan *et al.*, 2012) can be developed to eliminate observer effects and standardize detectability to provide counts with minimal (and quantifiable) bias. These extensions would further serve to build a more robust and automated observation model to combine with the flexible abundance model for the migration process described in this paper.

ACKNOWLEDGEMENTS

We are grateful to the observers who participated in field work, and for advice from Steve Reilly, Bill Perrin, Jeff Breiwick, Dave Rugh, Rod Hobbs and Jeff Laake. Thanks also to Paul Wade, Andre Punt and Jeff Laake for providing valuable comments on an earlier version of this manuscript.

REFERENCES

- Brooks, S.P. and Gelman, A. 1998. General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* 7: 434-455
- Buckland, S.T. and Breiwick, J.M. 2002. Estimated trends in abundance of eastern Pacific gray whales from shore counts, 1967/68 to 1995/96. *J. Cetacean Res. Manage.* 4:41-48.
- Buckland, S. T., Breiwick, J.M., Cattanach, K.L. and Laake, J.L. 1993. Estimated population size of the California gray whale. *Mar. Mammal Sci.* 9:235-249.
- Chelgren, N.D., Adams, M.J., Bailey, L.L. and Bury, R.B. 2011. Using multilevel spatial models to understand salamander site occupancy patterns after wildfire. *Ecology* 92: 408-421.
- Crainiceanu, C. M., Ruppert, D., and Wand, M. P. 2005. Bayesian analysis for penalized spline regression using WinBUGS. *J. Stat. Softw.* 14: 1-24.
- Durban, J., Lang, A., Weller, D., Rugh, D., Hobbs, R., and Perryman, W. 2010. Comparing shore-based counts of eastern North Pacific gray whales. Paper SC/62/BRG8 presented to the IWC Scientific Committee, May 2010 (unpublished). 6pp. [Available from the Office of this Journal].
- Durban, J., Weller, D., Lang, A., and Perryman, W. 2011. Abundance indices of eastern North Pacific gray whales from southbound migration counts, 2007-2011. Paper SC/63/BRG7 presented to the IWC Scientific Committee, May 2011 (unpublished). 4pp. [Available from the Office of this Journal].
- Gelman, A., Meng, X. L., and Stern, H. S. 1996. Posterior Predictive Assessment of Model Fitness via Realized Discrepancies. *Stat. Sinica* 6: 733-807.
- Hobbs, R.C., Rugh, D.J., Waite, J.M., Breiwick, J.M. and DeMaster, D.P. 2004. Abundance of North Pacific gray whales on the 1995/96 southbound migration. *J. Cetacean Res. Manage.* 6:115-120.
- Joseph, L., Elkin, C., Martin, T., and Possingham, H. 2009. Modeling abundance using N-mixture models: the importance of considering ecological mechanisms. *Ecol. Appl.* 19: 631-642
- Kery, M., Royle, J., and Schmid, H. 2005. Modeling avian abundance from replicated counts using binomial mixture models. *Ecol. Appl.* 15: 1450-1461.
- Laake, J.L., Rugh, D.J., Lerczak, J.A. and Buckland, S.T. 1994. Preliminary estimates of population size of gray whales from the 1992/93 and 1993/94 shore-based surveys. Paper SC/46/AS7 presented to the IWC Scientific Committee, May 1994 (unpublished). 13pp. [Available from the Office of this Journal].
- Laake, J.L., Punt, A.E., Hobbs, R., Ferguson, M., Rugh, D. and Breiwick, J. 2012. Gray whale southbound migration surveys 1967-2006: An integrated re-analysis. *J. Cetacean Res. Manage.* 12(3):287-306.
- Li, G., Best, N., Hansell, A.L., Ahmed, I. and Richardson, S. 2012, BayesSTDetect: detecting unusual temporal patterns in small area data via Bayesian model choice. *Biostatistics* doi:10.1093/biostatistics/kxs005 (Biostatistics Advance Access published March 26, 2012).
- Lunn, D.J., Thomas, A., Best, N. and Spiegelhalter, D. 2000. WinBUGS - a Bayesian modelling framework: concepts, structure and extensibility. *Stat. Comput.* 10: 325-337
- Kuo, L., Mallick, B. 1998. Variable selection for regression models. *Sankhya Ser. B* 60: 65-81.
- Perryman, W.L., Donahue, M.A., Laake, J.L. and Martin, T.E. 1999. Diel variation in migration rates of eastern Pacific gray whales measured with thermal imaging sensors. *Mar. Mammal Sci.* 15:426-445.
- Punt, A. E. and Wade, P.R. 2012. Population status of the eastern North Pacific stock of gray whales in 2009. *J. Cetacean Res. Manage.* 12:15-28.
- Royle, J. 2004. N-mixture models for estimating population size from spatially replicated counts. *Biometrics* 60: 108-115.
- Rugh, D.J., Shelden, K.E., and Schulman-Janiger, A. 2001. Timing of the gray whale southbound migration. *J. Cetacean Res. Manage.* 3, 31-40.
- Rugh, D.R., Hobbs, R.C., Lerczak, J.A., and Breiwick, J.M.. 2005. Estimates of abundance of the eastern North Pacific stock of gray whales (*Eschrichtius robustus*) 1997-2002. *J. Cetacean Res. Manage.* 7:1-12.
- Ruppert, D. 2002. Selecting the number of knots for penalized splines. *J. Comput. Graph. Stat.* 11: 735-757
- Santhaseelan, V., Arigela, S. and Asari, V.K. 2012. Neural Network Based Methodology for Automatic Detection of Whale Blows in Infrared Video. *Lect. Notes. Comput. Sc.* 7431: 230-240.